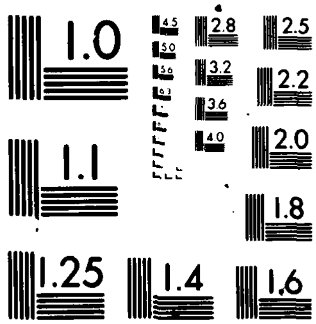
AD-A093 123 PITTSBURGH UNIV PA INST FOR STATISTICS AND APPLICATIONS F/G 12/1
ON THE CONVEXITY OF SOME DIVERGENCE MEASURES BASED ON ENTROPY F--ETC(U)
SEP 80 J BURBEA, C R RAO F49620-79-C-0161
UNCLASSIFIED TR-80-13 AFOSR-TR-80-1167 NL

1 of 1
AD A093123

END
DATE
FILMED
2-84
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0

AD A093123

ON THE CONVEXITY OF SOME DIVERGENCE MEASURES BASED ON ENTROPY FUNCTIONS.

J. Burbea

~~and~~

C. Radhakrishna Rao*

September 1980

Technical Report No. 80-13

Institute for Statistics and Applications
Department of Mathematics and Statistics
University of Pittsburgh
Pittsburgh, PA. 15260

*The work of this author is sponsored by the Air Force Office of Scientific Research under Contract [redacted]. Reproduction in whole or in part is permitted ~~for any purpose~~ of the United States Government.

F49620-79-C-0161

ON THE CONVEXITY OF SOME DIVERGENCE MEASURES BASED ON ENTROPY FUNCTIONS

J. Burbea and C. Radhakrishna Rao*

Abstract - Three measures of divergence between vectors in a convex set of an n-dimensional real vector space have been defined in terms of certain types of entropy functions, and their convexity property studied. Among other results, a classification of the α-order entropies is obtained by the convexity of these measures. These results have applications to the measurement of diversity of a discrete probability distribution and divergence between two distributions.

---

*Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania, 15260.

The work of the second author is sponsored by the Air Force Office of Scientific Research under Contract F 49620-79-0161.

AMS (MOS) Subject Classification: 94A17, 94A24 .

Key Words and Phrases: Cross entropy, Divergence, Entropy, Jensen difference.

1. INTRODUCTION

One of the most widely used index of diversity of a multinomial distribution, $x = (x_1, \ldots, x_n)$, $x_i \geq 0$, $\Sigma x_i = 1$, is the Shannon entropy, $H_n(x) = -\Sigma\, x_i \log x_i$ (Shannon [10]). The concavity of $H_n(x)$ provides a decomposition of the total diversity in a mixed distribution $(x+y)/2$ as

$$H_n(\tfrac{x+y}{2}) = \tfrac{1}{2}[H_n(x) + H_n(y)] + \tfrac{1}{2}\, J_n(x,y) \tag{1.1}$$

The first component $2^{-1}[H_n(x) + H_n(y)]$ in (1.1) is the average diversity within the distributions, and the second component

$$J_n(x,y) = [-H(x) - H(y)] - 2[-H\, \frac{(x+y)}{2}\,] \tag{1.2}$$

which we call the Jensen difference arising out of the convex function $-H(x)$ is non-negative, vanishes if and only if $x = y$, and thus provides a natural measure of divergence between the distributions x and y. (See Lewontin [6] and Rao [9] for some applications of $H_n(x)$ and $J_n(x,y)$ in biological studies). It is interesting to note that $J_n(x,y)$ considered as a function of (x,y) is convex, which meets the intuitive requirement that the average divergence between (x,y) and (z,w) is not less than that between their convex combination $\lambda(x,y) + \mu(z,w)$ where λ, $\mu \geq 0$ and $\lambda + \mu = 1$. The convexity of the divergence measure $J_n(x,y)$ is an additional attractive feature of the Shannon entropy $H_n(x)$ as a measure of diversity of a distribution.

In this paper we consider the Jensen difference (1.2) arising from a generalized class of entropy functions including the α-order entropies due to Havrda and Charvát [3], which we call the J-divergence and examine its convexity. In particular, we show that the J-divergence (1.2) based on the α-order entropy

$$H_{n,\alpha}(x) = (\alpha-1)^{-1} \, (1 - \Sigma \, x_i^{\alpha}) \, , \quad \alpha \neq 1 \tag{1.3}$$

defined on the convex set

$$S_n = \{(x_1, \ldots, x_n) \in I^n : \; \Sigma \, x_i = 1\}, \; I \equiv (0,1) \tag{1.4}$$

is convex on $S_n \times S_n$ if and only if $\alpha \in [1,2]$ for $n > 2$ and if and only if $\alpha \in [1,2]$ or $[3, 11/3]$ for $n = 2$. The last result is surprising and the proof is rather involved.

We define two other measures called the K and L-divergences (equations (2.4) and (2.5)) based on cross entropy functions (Good [2]) and study their convexity. These are similar to and include the divergence measure introduced by Jeffreys [4] for providing an invariant density of a priori probability and applied for the more general purpose of statistical inference by Kullback and Leibler [5].

As a by-product of these results we obtain some interesting inequalities (equations (4.3) and (5.7)).

We note that the J, K and L-divergences are semimetrics and not, in general, metrics as they may not satisfy the triangular inequality. However, by considering these

functions on a tangent space of a parametric space of probability distributions, one is led to a differential metric of a Riemannian geometry which induces a metric over the space of distribution functions. This was done earlier by Rao [7,8] where the differential metric is in terms of the information matrix of a parametric family of probability distributions. This metric has been recently studied by Atkinson and Mitchell [1]. Some extensions of this approach to more general convex functions along with other local properties of the J, K, L-divergences will be presented elsewhere. The present study is an investigation of the global properties of these divergence measures.

2. PRELIMINARIES AND NOTATION

Let ϕ be a C^2-function on a domain D of $\mathbb{R}^n$. The Hessian of ϕ at $x \in D$ along the direction $u \in \mathbb{R}^n$ is defined by

$$\Delta_u \phi(x) \equiv d^2 \phi(x:u) = u^T M_\phi u \ ,$$

where M_ϕ is the $n \times n$ matrix whose entries are $\partial_{x_i} \partial_{x_j} \phi(x)$; $i,j=1,\ldots,n$. This may also be written as

$$\Delta_u \phi(x) \equiv u^T [\partial_{x_i} \partial_{x_j} \phi] u \ .$$

Sometimes it is convenient to consider a function ψ as a function on the cartestan product in $R^n \times R^n$. In this case we assume that $\psi = \psi(\cdot,\cdot)$ is a C^2-function on $D \times D$. The

Hessian, then, of ψ at $(x,y) \in D\times D$ along the direction $(u,v) \in \mathbb{R}^n \times \mathbb{R}^n$ is given by

$$\Delta_{(u,v)}\psi(x,y) = u^T[\partial_{x_i}\partial_{x_j}\psi]u + 2v^T[\partial_{x_i}\partial_{y_j}\psi]u + v^T[\partial_{y_i}\partial_{y_j}\phi]v \quad (2.1)$$

with the obvious meaning of the expressions involved.

Let D be a convex domain of $\mathbb{R}^n$. A function ϕ of class $C^2(D)$ is said to be convex on D if for every $(x,u) \in D\times\mathbb{R}^n$, $\Delta_u\phi(x) \geq 0$. The smoothness assumption $\phi \in C^2(D)$ can be, of course, weakened by only requiring that ϕ be continuous on D with $\Delta_u\phi(x) \geq 0$, where the partial derivatives are taken in the distributional sense. Alternatively, one may apply a standard regularization process. We briefly recall this concept. We choose a C^∞-nonnegative function K whose compact support is inside the unit ball of $\mathbb{R}^n$ and such that

$$\int K(x)dx = 1.$$

For $\varepsilon > 0$ we define

$$K_\varepsilon(x) \equiv \varepsilon^{-n}K(\varepsilon^{-1}x).$$

Suppose f is locally integrable in the domain D of $\mathbb{R}^n$. We may assume that $f = 0$ outside a compact set and thus $f \in L_1(\mathbb{R}^n)$. We define

$$f_\varepsilon(y) \equiv (f * K_\varepsilon)(y) = \int f(x)K_\varepsilon(y-x)dx = \int K(x)f(y-\varepsilon x)dx .$$

As is well known, $f_\varepsilon \in C^\infty(D)$. Moreover, if in addition f is continuous on D, then it is uniformly continuous on compacta of D and, $\lim_{\varepsilon\to 0} f_\varepsilon = f$ uniformly on compacta of D.

For a function ϕ which is continuous on a convex domain D, but not necessarily of class $C^2(D)$, to be convex (in the generalized sense) in D, we may only require that its regularization ϕ_ε, defined above, be convex in D, in the previously described restrictive sense. It is said to be concave if $-\phi$ is convex. Thanks to the above process of regularization we may always assume that the functions in question are sufficiently smooth.

Let ϕ be a C^2-function on an interval I of $\mathbb{R}$ and consider the ϕ-entropy

$$H_{n,\phi}(x) = -\sum_{i=1}^{n} \phi(x_i), \quad x \in I^n \tag{2.2}$$

as a function defined on I^n. The Jensen difference (1.2) based on (2.2), which will be referred to as the J-divergence between x and y, is

$$J_{n,\phi}(x,y) = \sum_{i=1}^{n} \{\phi(x_i) + \phi(y_i) - 2\phi[(x_i+y_i)/2]\}, (x,y) \in I^n \times I^n. \tag{2.3}$$

When the interval I does not contain the origin, we consider alternative measures which may be called the K and L-divergences,

$$K_{n,\phi}(x,y) = \sum_{i=1}^{n} (x_i - y_i)\left[\frac{\phi(x_i)}{x_i} - \frac{\phi(y_i)}{y_i}\right] \tag{2.4}$$

and

$$L_{n,\phi}(x,y) = \sum_{i=1}^{n} \left[x_i \phi\left(\frac{y_i}{x_i}\right) + y_i \phi\left(\frac{x_i}{y_i}\right)\right] \tag{2.5}$$

The Hessians of (2.3)-(2.5) can be computed using the formula (2.1). However, it is of some practical interest to

consider the divergence measures (2.3)-(2.5) as acting on the convex set S_n defined in (1.4). In this case, (2.2) can be written as

$$H_{n,\phi}(x:X) = H_{n-1,\phi}(x) + H_{1,\phi}(X) \tag{2.6}$$

$$x = (x_1,\ldots,x_{n-1}) \in I^{n-1}, \quad X = 1 - \sum_{i=1}^{n-1} x_i \in I. \tag{2.7}$$

Then (2.3) may be written as

$$J_{n,\phi}(x:X,y:Y) = J_{n-1,\phi}(x,y) + J_{1,\phi}(X,Y) \tag{2.8}$$

where y,Y are defined in the same way as x,X. Similar expressions for the K and L-divergences (2.4) and (2.5) are also available.

Note that

$$\Delta_u H_{n,\phi}(x:X) = \Delta_u H_{n-1,\phi}(x) + \Delta_U H_{1,\phi}(X) \tag{2.9}$$

and the Hessian of (2.3) subject to (2.7) is

$$\Delta_{u,v} J_{n,\phi}(x:X,\ y:Y) = \Delta_{u,v} J_{n-1,\phi}(x,y) + \Delta_{U,V} J_{1,\phi}(X,Y) \tag{2.10}$$

with similar expressions for the K and L-divergences, where

$$u = (u_1,\ldots,u_{n-1}),\ v = (v_1,\ldots,v_{n-1}) \in \mathbb{R}^{n-1}$$

and

$$U = \sum_{i=1}^{n-1} u_i \ , \qquad V = \sum_{i=1}^{n-1} v_i \in \mathbb{R}.$$

We denote by

$$\bar{S}_n = \{(x_1,\ldots,x_n) \in \bar{I}^n:\ \Sigma\, x_i = 1\};\ \bar{I} = [0,1],\ n \geq 2,$$

the closure of S_n defined in (1.4). For any real number α, we define

$$\phi_\alpha(x) = \begin{cases} (\alpha-1)^{-1}(x^\alpha - x), & \alpha \neq 1 \\ x \log x, & \alpha = 1 \end{cases} \tag{2.11}$$

over $x \in \mathbb{R}_+ \equiv (0,\infty)$, and when $\alpha \geq 0$, ϕ_α can be extended to $x = 0$ with the convention $0 \log 0 = 0$. Defining

$$H_{n,\alpha}(x) \equiv H_{n,\phi_\alpha}(x), \quad x \in S_n \tag{2.12}$$

we have

$$H_{n,1}(x) = -\Sigma\, x_i \log x_i, \qquad x \in \bar{S}_n, \tag{2.13}$$

$$H_{n,\alpha}(x) = (\alpha-1)^{-1}(1 - \Sigma\, x_i^\alpha), \quad x \in S_n, \ \alpha \neq 1. \tag{2.14}$$

We note that $H_{n,\alpha}$, for $\alpha \geq 0$ can be extended to the closure $\bar{S}_n$, which is the α-order entropy introduced by Havrda and Charvát [3], and that $H_{n,\alpha}$ tends to $H_{n,1}$ as $\alpha \to 1$, which is the Shannon entropy H_n.

The J,K and L-divergences based on $H_{n,\alpha}$ are denoted by $J_{n,\alpha}$, $K_{n,\alpha}$ and $L_{n,\alpha}$ respectively. Their explicit expressions are as follows:

$$J_{n,\alpha}(x,y) = \begin{cases} (\alpha-1)^{-1}\, \Sigma\{x_i^\alpha + y_i^\alpha - 2[(x_i+y_i/2]^\alpha\}, & \alpha \neq 1 \\ \Sigma\{x_i \log x_i + y_i \log y_i - (x_i+y_i) \log[(x_i+y_i)/2]\}, & \alpha = 1 \end{cases} \tag{2.15}$$

$$K_{n,\alpha}(x,y) = \begin{cases} (\alpha-1)^{-1} \Sigma(x_i-y_i)(x_i^{\alpha-1} - y_i^{\alpha-1}), & \alpha \neq 1 \\ \Sigma(x_i-y_i)(\log x_i - \log y_i), & \alpha=1 \end{cases} \tag{2.16}$$

and

$$L_{n,\alpha}(x,y) = \begin{cases} (\alpha-1)^{-1}\{\Sigma x_i^{\alpha} y_i^{1-\alpha} + \Sigma x_i^{1-\alpha} y_i^{\alpha} - 2\}, & \alpha \neq 1 \\ \Sigma(x_i-y_i)(\log x_i - \log y_i), & \alpha=1 . \end{cases} \tag{2.17}$$

Here $(x,y) \in S_n \times S_n$, and for $\alpha \geq 0$, $J_{n,\alpha}$ can be extended to $\bar{S}_n \times \bar{S}_n$. We note that $K_{n,1} = L_{n,1}$, and these expressions are the same as the divergence measure of Jeffreys [4] and Kullback and Liebler [5].

3. THE J-DIVERGENCE

The Hessian of $J_{n,\phi}$, in view of (2.1), is given by

$$\Delta_{(u,v)} J_{n,\phi}(x,y) = \sum_{i=1}^{n} \{a(x_i,y_i)u_i^2 + 2b(x_i,y_i)u_i v_i + a(y_i,x_i)v_i^2\} \tag{3.1}$$

where $x,y \in I^n$ with I being any interval of the line. Here, for $x,y \in I$,

$$b(x,y) = -\frac{1}{2}\, \phi''[(x+y)/2] \tag{3.2}$$

and

$$a(x,y) = \phi''(x)+b(x,y) \; ; \; x,y \in I . \tag{3.3}$$

This shows that $J_{n,\phi}$ is convex (concave) on $I^n \times I^n$ if and only if $a(x,y) \geq 0$ (or $a(x,y) \leq 0$) and

$$d(x,y) \equiv a(x,y)a(y,x) - [b(x,y)]^2 \geq 0 \tag{3.4}$$

for every $(x,y) \in I \times I$.

Now, using (3.2)-(3.4) we deduce that for $x,y \in I$,

$$a(x,y) = \phi''(x)\,\phi''[(x+y)/2]\left\{\frac{1}{\phi''[(x+y)/2]} - \frac{1}{2}\frac{1}{\phi''(x)}\right\}$$

and

$$d(x,y) = \phi''(x)\,\phi''(y)\,\phi''[(x+y)/2]$$
$$\times\left\{\frac{1}{\phi''[(x+y)/2]} - \frac{1}{2\phi''(x)} - \frac{1}{2\phi''(y)}\right\}.$$

The expression in the last curly bracket is directly related to the Jensen difference of $(\phi'')^{-1}$. This with a closer examination of these expressions leads to the following basic result:

Theorem 1. $J_{n,\phi}$ is convex (concave) on $I^n \times I^n$ if and only if ϕ is convex (concave) and $(\phi'')^{-1}$ is concave (convex) on I.

As an application of the theorem we consider the following family of functions

$$g_\alpha(x) = a\,f_\alpha(x) + bx + c \tag{3.5}$$

where a,b,c are arbitrary constants and $\{f_\alpha\}$ is a one parameter family of nonnegative functions defined on an interval I such that

$$f_\alpha''(x) = \alpha(\alpha-1)f_{\alpha-2}(x) \; ; \; x \in I, \; \alpha \in \mathbb{R} . \tag{3.6}$$

We shall fix a normalization

$$a\alpha(\alpha-1) \geq 0, \tag{3.7}$$

from which it follows that g_α is convex on I for any $\alpha \in \mathbb{R}$. An immediate consequence of Theorem 1 is the following:

Corollary 1. Let the notation of (3.5)-(3.7) apply and consider H_{n,g_α} and J_{n,g_α} as formed in (2.2)-(2.3). Then, for any $\alpha \in \mathbb{R}$, H_{n,g_α} is concave on I^n while J_{n,g_α} is never concave on $I^n \times I^n$. Moreover, J_{n,g_α} is convex on $I^n \times I^n$ if and only if $(f_{\alpha-2})^{-1}$ is concave on I.

This corollary is appled to the following special case

$$f_\alpha(x) = x^\alpha \quad , \quad x \in \mathbb{R}_+ .$$

Writing $\beta=\alpha-2$, we examine whether $h_\beta \equiv (f_\beta)^{-1}$ is concave on $\mathbb{R}_+$. We have

$$h_\beta''(x) = \beta(\beta-1)x^{-\beta-2} \quad , \quad x \in \mathbb{R}_+$$

and thus, h_β is concave if and only if $\beta \in [-1,0]$. This yields the following result:

Corollary 2. Let

$$g_\alpha(x) = ax^\alpha + bx + c, \; x \in \mathbb{R}_+$$

where a,b,c and α are constants with $a\alpha(\alpha-1) \geq 0$. Then H_{n,g_α} is concave on $\mathbb{R}_+^n$ while J_{n,g_α} is never concave on $\mathbb{R}_+^n \times \mathbb{R}_+^n$.

Moreover, J_{n,g_α} is convex on $\mathbb{R}_+^n \times \mathbb{R}_+^n$ if and only if $\alpha \in [1,2]$ in which case $a \geq 0$.

Instead of g_α in this corollary we may take ϕ_α as in (2.11), and consequently:

Corollary 3. For any $\alpha \geq 0$, H_{n,ϕ_α} is concave on $\mathbb{R}_+^n$ and J_{n,ϕ_α} is never concave on $\mathbb{R}_+^n \times \mathbb{R}_+^n$. Moreover, J_{n,ϕ_α} is convex on $\mathbb{R}_+^n \times \mathbb{R}_+^n$ if and only if $\alpha \in [1,2]$.

Using this corollary and (2.6)-(2.10) we see that $J_{n,\alpha}$, for $n \geq 3$, is convex on $\bar{S}_n \times \bar{S}_n$ if and only if $\alpha \in [1,2]$. Of course, $J_{2,\alpha}$ is also convex on $\bar{S}_2 \times \bar{S}_2$ for every $\alpha \in [1,2]$. However, $J_{2,\alpha}$, interestingly, is also convex for other values of α, viz., in $[3,11/3]$. The proof of this fact is postponed to the next section. Meanwhile, we shall record the following corollary:

Corollary 4. For any $\alpha \geq 0$, $H_{n,\alpha}$ of (2.12) is concave on $\bar{S}_n$ and $J_{n,\alpha}$ of (2.15) is never concave on $\bar{S}_n \times \bar{S}_n$. Moreover, for $n \geq 3$, $J_{n,\alpha}$ is convex on $\bar{S}_n \times \bar{S}_n$ if and only if $\alpha \in [1,2]$. Also, if $\alpha \in [1,2]$ then $J_{2,\alpha}$ is convex on $\bar{S}_2 \times \bar{S}_2$.

4. ADDITIONAL PROPERTIES OF THE J-DIVERGENCE

In order to deal with $J_{2,\alpha}$ on $\bar{S}_2 \times \bar{S}_2$ we shall apply Corollary 1 to the following family

$$f_\alpha(x) = x^\alpha + (1-x)^\alpha \quad ; \quad x \in I \equiv [0,1].$$

For this purpose we shall establish the following Lemma which is of some interest on its own right.

Lemma 1. The function

$$h_\beta(x) \equiv [f_\beta(x)]^{-1} = \{x^\beta+(1-x)^\beta\}^{-1} \quad ; \ x \in I = [0,1],$$

has the following properties:

(i) for $\beta \in (-\infty,-1)$ and $\beta \in [2,\infty)$, h_β has inflection points on I;

(ii) for $\beta \in (0,1)$, h_β is (strictly) convex on I;

(iii) for $\beta \in [-1,0]$, h_β is concave on I;

(iv) for $\beta \in [1,5/3]$, h_β is concave on I while for $\beta \in (5/3,2)$, h_β has inflection points on I.

Proof. We have

$$h_\beta'' = f_\beta^{-3}[2(f_\beta')^2 - \beta(\beta-1)f_\beta f_{\beta-2}]$$

and item (ii) follows at once. To proceed with the other items, we study the sign of the function

$$\begin{aligned} &2(f_\beta')^2 - \beta(\beta-1)f_\beta f_{\beta-2} \\ &= 2\beta^2[x^{\beta-1}-(1-x)^{\beta-1}]^2 - \beta(\beta-1)[x^{\beta-2}+(1-x)^{\beta-2}][x^\beta+(1-x)^\beta] . \end{aligned}$$

This function is symmetric about the point $x=1/2$ and it is therefore more convenient to introduce the new variable, $y=(1-x)/x$ with $y \in [0,1]$. This corresponds to $x \in [1/2,1]$ and by symmetry y may also be allowed to range in $[1,\infty]$. With this new

variable, the sign of the above function is the same as that of

$$F_\beta(y) \equiv \beta\{2\beta(1-y^{\beta-1})^2 - (\beta-1)(1+y^\beta)(1+y^{\beta-2})\} .$$

This may be also written as

$$F_\beta(y) = \beta\{(\beta+1)(1-y^{\beta-1})^2 - (\beta-1)y^{\beta-2}(1+y)^2\}. \tag{4.1}$$

When $\beta \in [-1,0]$ it follows from (4.1) that $F_\beta(y) \leq 0$ and therefore item (iii) follows. As for item (i), we see from (4.1) that

$$F_\beta(0) = +\infty \quad , \quad F_\beta(1) = 4\beta(1-\beta) < 0 \quad \text{for } \beta \in (-\infty,-1),$$

$$F_2(0) = 4 \quad , \quad F_2(1) = -8$$

and

$$F_\beta(0) = \beta(\beta+1) > 0, \quad F_\beta(1) = -4\beta(\beta-1) < 0, \text{ for } \beta \in (2,\infty) .$$

Consequently, item (i) follows. We turn now to item (iv). Here $F_1(y) \equiv 0$ and we shall therefore assume that $\beta \in (1,2)$. A differentiation of (4.1) gives

$$F_\beta'(y) = \beta(\beta-1)y^{\beta-3}\{2(\beta+1)y^\beta - \beta y^2 - 4\beta y + 2 - \beta\}.$$

The sign of this derivative is determined by

$$G_\beta(y) \equiv 2(\beta+1)y^\beta - \beta y^2 - 4\beta y + 2 - \beta .$$

Now,

$$G_\beta(0) = 2 - \beta > 0 \quad , \quad G_\beta(1) = -4(\beta-1) < 0 ,$$

and hence $G_\beta(y_\beta) = 0$ for some $y_\beta \in (0,1)$. Next, we have

$$G_\beta'(y) = 2\beta\{(\beta+1)y^{\beta-1} - y - 2\} .$$

However, by Bernoulli's inequality

$$\begin{aligned} y + 2 - (\beta+1)y^{\beta-1} &= y + 2 - (\beta+1)[1-(1-y)]^{\beta-1} \\ &\geq y + 2 - (\beta+1)[1-(\beta-1)(1-y)] \\ &= (2-\beta^2)y + \beta(\beta-1). \end{aligned}$$

The last expression describes a straight line passing through the points $(0,\beta^2-\beta)$ and $(1,2-\beta)$ and therefore

$$y+2 - (\beta+1)y^{\beta-1} > 0 \quad \text{for} \quad y \in (0,1)$$

Consequently, y_β is the only root of $G_\beta(y) = 0$ in $(0,1)$ and, moreover, $F_\beta(y)$ has a <u>single maximum</u> at $y_\beta \in (0,1)$. The root y_β lies in the variety.

$$2(\beta+1)y^\beta = \beta y^2 + 4\beta y + \beta-2 . \tag{4.2}$$

We replace y^β in (4.1) by the quadratic expression in (4.2). This, after some manipulations, results in

$$H_\beta(y) \equiv -4 \frac{\beta+1}{\beta^2} y^2 F_\beta(y) = (\beta-2)y^4+8(\beta-1)y^3+2(7\beta-6)y^2+8(\beta-1)y+\beta-2$$

and, hence, we seek β for which $H_\beta(y_\beta) \geq 0$. However, we can factor $H_\beta(y)$ in the form of

$$H_\beta(y) = (\beta-2)(1+y)^2 [y-B(\beta)][y-B(\beta)^{-1}]$$

where

$$B(\beta) \equiv (2-\beta)^{-1}\{3\beta-2-2[2\beta(\beta-1)]^{\frac{1}{2}}\}.$$

Since $\beta \in (1,2)$, we clearly have $0 < B(\beta) < 1 < B(\beta)^{-1}$. Hence, $H_\beta(y_\beta) \geq 0$ if and only if

$$y_\beta \geq B(\beta).$$

This condition is equivalent to the requirement that $F_\beta[B(\beta)] \geq 0$. This requirement is determined by the region of non-negativity of the function $K(\beta)$ defined below. This interesting function is defined as follows:

$$K(\beta) \equiv 2(\beta+1)B(\beta)^{\beta} - \beta B(\beta)^{2} - 4\beta B(\beta) + 2 - \beta \quad ; \quad \beta \in (1,2).$$

We have

$$K(1)=K(2)=0 \quad ; \quad K'(1)=+\infty \quad , \quad K'(2)=0.$$

Moreover, a direct calculation shows that $K(5/3)=0$ and that $\beta = 5/3$ is the cut-off point of the region of non-negativity. Thus $K(\beta) > 0$ for all $\beta \in (1,5/3)$, $K(5/3) = 0$ and $K(\beta) < 0$ for all $\beta \in (5/3,2)$, (see Figure 1). The proof of the lemma is now complete.

Before proceeding any further we shall record an interesting consequence of this lemma, or rather from the proof of the lemma.

Corollary 5. For any $\gamma \in [0,2/3]$ the following inequality holds for all $t \in (-\infty,\infty)$

$$\left(\frac{\sinh \gamma t}{\cosh t}\right)^{2} \leq \frac{\gamma}{\gamma+2} \quad . \tag{4.3}$$

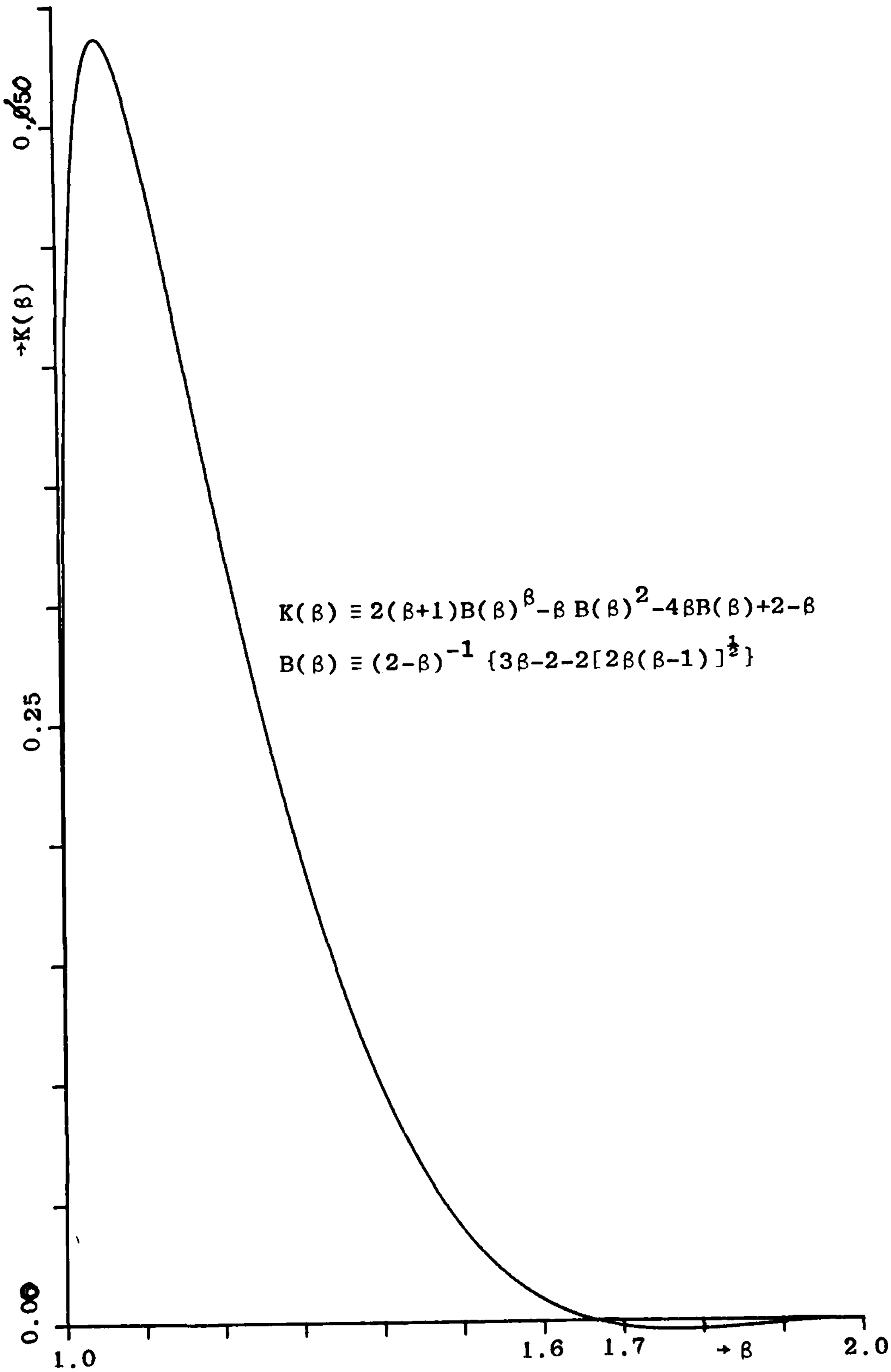

0.50
→K(β)
0.25
0.00
$K(\beta) \equiv 2(\beta+1)B(\beta)^{\beta}-\beta\, B(\beta)^{2}-4\beta B(\beta)+2-\beta$
$B(\beta) \equiv (2-\beta)^{-1}\,\{3\beta-2-2[2\beta(\beta-1)]^{\frac{1}{2}}\}$
1.0
1.6
1.7
→β
2.0

Figure 1

Proof. From (4.1) we know that $F_\beta(y) \leq 0$ for $\beta \in [1,5/3]$ and, therefore,

$$(\beta+1)(1-y^{\beta-1})^2 \leq (\beta-1)y^{\beta-2}(1+y)^2.$$

This was proved for $y \in [0,1]$. However, this inequality is invariant under the substitution $y \to y^{-1}$ and, therefore, it is valid for all $y \in (0,\infty)$. Setting $y^{\frac{1}{2}} = e^t$ and $\beta = \gamma + 1$ concludes the proof.

Corresponding to Theorem 1 and Corollaries 1 and 2, Lemma 1 leads to :

Theorem 2. Let

$$g_\alpha(x) = af_\alpha(x) + bx + c \quad , \quad x \in I = [0,1] \; ,$$

where a,b,c and α are constants with $a\alpha(\alpha-1) \geq 0$ and

$$f_\alpha(x) = x^\alpha + (1-x)^\alpha \; .$$

Then H_{n,g_α} is convex on I^n. Moreover, J_{n,g_α} is never concave on $I^n \times I^n$. It is convex there if and only if $\alpha \in [1,2]$ or $\alpha \in [3,11/3]$, in which case $a \geq 0$.

Theorem 2 enables us to strengthen the result of Corollary 4 on the Jensen difference of the α-order entropy with the following additional feature:

Corollary 6. $J_{2,\alpha}$ is convex on $\bar{S}_2 \times \bar{S}_2$ if and only if $\alpha \in [1,2]$ or $\alpha \in [3,11/3]$.

In correspondence with (2.11) we define

$$g_\alpha(x) = \begin{cases} (\alpha-1)^{-1}[x^\alpha + (1-x)^\alpha] & , \ \alpha \neq 1 \\ x \log x + (1-x)\log(1-x) & , \ \alpha = 1 \end{cases} \tag{4.4}$$

for $x \in I \equiv [0,1]$. We also define

$$G_{n,\alpha}(x) \equiv H_{n,g_\alpha}(x) \quad , \quad x \in S_n \, , \tag{4.5}$$

and call $G_{n,\alpha}(x)$, $x \in S_n$, the <u>paired entropy of order α</u>. Using (2.11) - (2.14), we clearly have the following relationships:

$$G_{n,\alpha}(x) = H_{n,\alpha}(x) + H_{n,\alpha}(1-x) - (\alpha-1)^{-1} \; ; \; \alpha \neq 1 \, , \quad x \in \bar{S}_n$$

$$G_{n,1}(x) = H_{n,1}(x) + H_{n,1}(1-x) \quad ; \quad x \in \bar{S}_n$$

We shall write

$$I_{n,\alpha}(x,y) \equiv J_{n,g_\alpha}(x,y) \; ; \; (x,y) \in S_n \times S_n \tag{4.6}$$

for the Jensen difference of g_α of (4.4). From Theorems 1, 2 and (2.6)-(2.10) we conclude :

<u>Theorem 3</u>. Let the notation of (4.4)-(4.6) apply with $\alpha \geq 0$. Then:

(i) $G_{n,\alpha}$ is concave on $\bar{S}_n$;

(ii) $I_{n,\alpha}$ is never concave on $\bar{S}_n \times \bar{S}_n$;

(iii) $I_{n,\alpha}$ is convex on $\bar{S}_n \times \bar{S}_n$ if and only if $\alpha \in [1,2]$ or $\alpha \in [3,11/3]$.

In particular,

(iv) $G_{n,1}$ is concave on $\bar{S}_n$ and $I_{n,1}$ is convex on $\bar{S}_n \times \bar{S}_n$.

Item (iv) of this theorem is a limiting case of the previous items as $\alpha \to 1$. It could also be directly deduced from Theorem 1. Indeed, from (4.4), $g_1''(x) = [x(1-x)]^{-1} > 0$ which

shows that g_1 is convex on (o,1). Furthermore, $F=(g_1'')^{-1}$ is given by $F(x)=x-x^2$ and thus $F''(x)=-2<0$. Therefore, $(g_1'')^{-1}$ is concave on [0,1] and Theorem 1 applies.

It may be noted that we could base our analysis of sections 3 and 4 on a more generalized form of the Jensen difference

$$J_{\phi}^{(\alpha,\beta)}(x,y)=2[\alpha\,\phi(x)+\beta\,\phi(y)-\phi(\alpha x+\beta y)] \tag{4.7}$$

with $\alpha,\beta\geq 0$, $\alpha+\beta=1$, so that (4.7) reduces to J_{ϕ} when $\alpha=\beta$. However, this does not constitute a major generalization and the results obtained for J_{ϕ} can also be derived for $J_{\phi}^{(\alpha,\beta)}$ after a minor modification of the argument.

5. THE K-DIVERGENCE

We briefly discuss the K-divergence $K_{n,\phi}$ defined in (2.4) and its relationship with the J-divergence $J_{n,\phi}$. To do this we define

$$\psi(x)\equiv\phi(x)/x\ ;\ x\in\mathbb{R}_+\ . \tag{5.1}$$

We start with the following simple proposition:

<u>Proposition 1</u>. $K_{n,\phi}$ is non-negative on $\mathbb{R}_+^n\times\mathbb{R}_+^n$ if and only if ψ is increasing on $\mathbb{R}_+$.

<u>Proof</u>. This is equivalent to the specialized statement with

$n=1$ which in turn is straightforward.

The following theorem establishes a comparison between $K_{n,\phi}$ and $J_{n,\phi}$:

Theorem 4. Assume that ψ is increasing and concave on I. Then, for any $(x,y) \in \mathbb{R}_+^n \times \mathbb{R}_+^n$,

$$J_{n,\phi}(x,y) \le K_{n,\phi}(x,y)$$

with equality if and only if $x = y$.

Proof. Again, this statement is equivalent to the specialized case of $n=1$. Accordingly, we consider the function

$$F(x,y) \equiv J_{1,\phi}(x,y)-K_{1,\phi}(x,y) \quad ; \quad (x,y) \in \mathbb{R}_+ \times \mathbb{R}_+ .$$

This may be written as

$$\frac{F(x,y)}{x+y} = \frac{y}{x+y}\, \psi(x) + \frac{x}{x+y}\, \psi(y) - \psi[(x+y)/2]$$

$$\le \psi\left(\frac{yx}{x+y} + \frac{xy}{x+y}\right) - \psi\left(\frac{x+y}{2}\right)$$

$$= \psi\left(\frac{2xy}{x+y}\right) - \psi\left(\frac{x+y}{2}\right) \le 0.$$

The first inequality follows from the concavity of ψ while the second inequality is due to the fact that ψ is increasing on $\mathbb{R}_+$. The equality statement also follows and the proof is complete.

The Hessian of $K_{n,\phi}$, in accordance with (2.1), is given

by

$$\Delta_{(u,v)}K_{n,\phi}(x,y) = \sum_{i=1}^{n} \{a(x_i,y_i)u_i^2 + 2b(x_i,y_i)u_iv_i + a(y_i,x_i)v_i^2\} \tag{5.2}$$

where $x,y \in \mathbb{R}_+^n$ and for $x,y \in \mathbb{R}_+$,

$$a(x,y) = \phi''(x) - y\,\psi''(x) \tag{5.3}$$

and

$$b(x,y) = -[\psi'(x)+\psi'(y)] \tag{5.4}$$

with ψ as given in (5.1). It follows, therefore, that $K_{n,\phi}$ is convex if and only if $a(x,y) \geq 0$ and

$$d(x,y) \equiv a(x,y)a(y,x) - [b(x,y)]^2 \geq 0 \quad ; \quad x,y \in \mathbb{R}_+ . \tag{5.5}$$

From (5.3) we see that $a(x,y) \geq 0$ whenever ϕ is convex and ψ is concave on $\mathbb{R}_+$. We have:

<u>Theorem 5</u>. Assume that ϕ is convex and ψ is concave on $\mathbb{R}_+$. Then:

(i) ψ is increasing on $\mathbb{R}_+$;

(ii) $K_{n,\phi}(x,y) \geq J_{n,\phi}(x,y) \geq 0$ for every $(x,y) \in \mathbb{R}_+^n \times \mathbb{R}_+^n$. Equality in one of the inequalities entails equalities in both inequalities. This occurs if and only if $x=y$.

If, in addition, (5.5) holds, then:

(iii) $K_{n,\phi}$ is convex on $\mathbb{R}_+^n \times \mathbb{R}_+^n$;

(iv) $K_{n,\phi}$ is convex on $S_n \times S_n$.

<u>Proof</u>. Using (5.1) we have

$$\psi'(x) = -\frac{1}{x}\,[\psi(x)-\phi'(x)]$$

and thus

$$\psi''(x) = -\frac{1}{x}[\psi'(x)-\phi''(x)] + \frac{1}{x^2}[\psi(x)-\phi'(x)]$$

$$= -\frac{1}{x}[2\psi'(x)-\phi''(x)] .$$

Therefore

$$2\psi'(x) = -x\psi''(x)+\phi''(x) \geq 0$$

and (i) follows. The fact that $J_{n,\phi}(x,y) \geq 0$ and its equality statement is a result of ϕ being convex. Also, $K_{n,\phi}(x,y) \geq J_n(x,y)$ and its equality statement follows from item (i) because of Porposition 1 and Theorem 4. This proves item (ii). Item (iii) follows from item (i) and the preceeding discussion. Item (iv) follows from (iii), (5.2) and formulae similar to (2.6) - (2.10). This concludes the proof.

The following hold:

Theorem 6. Let $\alpha \in [1,2]$. Then:

(i) $K_{n,\phi_\alpha}(x,y) \geq J_{n,\phi_\alpha}(x,y) \geq 0$ for every $(x,y) \in \mathbb{R}_+^n \times \mathbb{R}_+^n$. Equality in one of the inequalities occurs if and only if x=y. The same applies to $K_{n,\alpha}(x,y) \geq J_{n,\alpha}(x,y) \geq 0$ for every $(x,y) \in S_n \times S_n$.

(ii) K_{n,ϕ_α} is convex on $\mathbb{R}_+^n \times \mathbb{R}_+^n$ and $K_{n,\alpha}$ is convex on $S_n \times S_n$.

Proof. In this case ϕ_α is convex and ψ_α is concave on $\mathbb{R}_+$ and, therefore, we may use Theorem 5. To do so, we have to validate (5.5), i.e., we have to show that the discriminant function

$$d_\alpha(x,y) \equiv [\alpha x^{\alpha-2} - (\alpha-2)yx^{\alpha-3}][\alpha y^{\alpha-2}-(\alpha-2)xy^{\alpha-3}]-(x^{\alpha-2}+y^{\alpha-2})^2$$

is non-negative on $\mathbb{R}_+ \times \mathbb{R}_+$. Here, $d_1(x,y) \equiv d_2(x,y) \equiv 0$; we may, therefore, assume that $\alpha \in (1,2)$. Since $d_\alpha(x,x)=0$ and $d_\alpha(x,y)=d_\alpha(y,x)$ it is sufficient to assume that $y > x > 0$. In this way, we have

$$d_\alpha(x,y) = x^{2\alpha-4}\, f_\alpha(t) \;\; ; \;\; t \in y/x,$$

where

$$f_\alpha(t) \equiv t^{\alpha-3}[\alpha t-(\alpha-2)][\alpha-(\alpha-2)t] - (1+t^{\alpha-2})^2 \; . \tag{5.6}$$

We must show that $f_\alpha(t) \geq 0$ for $t \in (1,\infty)$. After some simplifications, we obtain

$$f'_\alpha(t) = (2-\alpha)t^{\alpha-4}g_\alpha(t)$$

with

$$g_\alpha(t) \equiv \alpha(\alpha-1)t^2 - 2(\alpha-1)^2 t - \alpha(3-\alpha) + 2t^{\alpha-1} \; .$$

Therefore,

$$g'_\alpha(t) = 2(\alpha-1)[\alpha(t-1)+1+t^{\alpha-2}] > 0 \;\; ; \;\; t \in (1,\infty),\ \alpha \in (1,2) \; .$$

Hence g_α is increasing on $(1,\infty)$ and since $g_\alpha(1) = 0$, we conclude that $g_\alpha(t) > 0$. Therefore, $f'_\alpha(t) > 0$ or that f_α is increasing on $(0,\infty)$. However, $f_\alpha(1) = 0$ and thus $f_\alpha(t) > 0$ for $t \in (1,\infty)$. This concludes the proof.

From the proof of this theorem we also deduce the following inequality:

<u>Corollary 6</u>. Let $\beta \in [0,1/2]$ Then , for every $s \in (-\infty,\infty)$,

$$\cosh^2\beta s \le [\beta^2 + (1-\beta^2)][1 + \frac{2\beta(1-\beta)}{\beta^2+(1-\beta)^2}\cosh s]. \tag{5.7}$$

Proof. For $\alpha \in [1,2]$ we have shown that f_α of (5.6) satisfies $f_\alpha(t) \ge 0$ for every $t \in [1,\infty)$. This is equivalent to

$$[\alpha t - (\alpha-2)][\alpha t^{-1} - (\alpha-2)] \ge [t^{(2-\alpha)/2} + t^{-(2-\alpha)/2}]$$

for every $t \in [1,\infty)$. Since this inequality is invariant under the transition $t \to t^{-1}$, it holds for every $t \in (0,\infty)$. Putting $t = e^s$ and $\beta = (2-\alpha)/2$ concludes the proof.

6. THE L-DIVERGENCE

The Hessian of $L_{n,\phi}(x,y)$ defined in (2.5), in view of (2.1), is

$$\Delta_{(u,v)} L_{n,\phi}(x,y) = \sum_{i=1}^{n} \{a(x_i,y_i)u_i^2 + 2b(x_i,y_i)u_iv_i + a(y_i,x_i)v_i^2\}$$

where $(x,y) \in \mathbb{R}_+^n \times \mathbb{R}_+^n$. Here

$$a(x,y) = \frac{1}{y}\phi''(\frac{x}{y}) + (\frac{y^2}{x^2}\phi''(\frac{y}{x})$$

and

$$b(x,y) = -\frac{x}{y^2}\phi''(\frac{x}{y}) - \frac{y}{x^2}\phi''(\frac{y}{x}) \ ; \ x,y \in \mathbb{R}_+$$

In this case, the discriminant

$$d(x,y) = a(x,y)a(y,x) - [b(x,y)]^2$$

is identically zero on $\mathbb{R}_+ \times \mathbb{R}_+$. This, together with formulae similar to (2.6)-(2.10), leads to:

Theorem 7. The following hold:

(i) $L_{n,\phi}(x,y) \geq 0$ for every $n \geq 1$ and every $(x,y) \in \mathbb{R}_+^n \times \mathbb{R}_+^n$ if and only if the function $\psi(t) \equiv t\,\phi(t^{-1}) + \phi(t)$ is non-negative for all $t \in \mathbb{R}_+$;

(ii) $L_{n,\phi}$ is convex on $\mathbb{R}_+^n \times \mathbb{R}_+^n$ if and only if $\psi(t) \equiv t\,\phi(t^{-1}) + \phi(t)$ is convex on $\mathbb{R}_+$.

Proof. As for item (i), we have

$$L_{n,\phi}(x,y) = \sum_{i=1}^{n} \frac{1}{x_i}\, \psi(t_i) \quad ; \quad t_i = y_i/x_i$$

and $L_{1,\phi}(x,y) = x^{-1}\psi(t)$, $t = y/x$. Thus (i) follows. As for item (ii), since $d(x,y) \equiv 0$ for every $(x,y) \in \mathbb{R}_+ \times \mathbb{R}_+$ we have that $D_{n,\phi}$ is convex on $\mathbb{R}_+^n \times \mathbb{R}_+^n$ if and only if

$$a(x,y) = \frac{y^2}{x^3}\left\{\frac{x^3}{y^3}\,\phi''\left(\frac{x}{y}\right) + \phi''\left(\frac{y}{x}\right)\right\} \geq 0 \quad ; \quad (x,y) \in \mathbb{R}_+ \times \mathbb{R}_+ .$$

Putting $t = y/x$ this condition becomes

$$t^{-3}\phi''(t^{-1}) + \phi''(t) \geq 0 \quad ; \quad t \in \mathbb{R}_+ .$$

This means that $\psi''(t) \geq 0$ and the theorem follows.

Corollary 7. For any $\alpha \geq 0$, $L_{n,\alpha}$ is a non-negative convex function on $S_n \times S_n$.

Proof. We use Theorem 7 and formulae similar to (2.6)-(2.10) for $\Delta_{(u,v)} L_{n,\alpha}(x,y)$ on $S_n \times S_n$. We start with

$\alpha = 1$. In this case

$$\phi_1(t) = t \log t \quad , \quad \psi_1(t) \equiv t\phi_1(t^{-1}) + \phi_1(t) \quad ; \quad t \in \mathbb{R}_+ ,$$

and thus

$$\psi_1(t) = (t-1)\log t \geq 0 \, , \quad \psi_1''(t) = (t^{-1} + t^{-2}) > 0 \, , \quad t \in \mathbb{R}_+ .$$

On the other hand, for $\alpha \neq 1$,

$$\phi_\alpha(t) = (\alpha-1)^{-1}(t^\alpha - t), \quad \psi_\alpha(t) \equiv t\phi_\alpha(t^{-1}) + \phi_\alpha(t) \; ; \; t \in \mathbb{R}_+ .$$

Therefore, for $\alpha \geq 0$, $\alpha \neq 1$,

$$\psi_\alpha(t) = (\alpha-1)^{-1}\, t^{1-\alpha}\, (t^{\alpha-1} - 1)\,(t^\alpha - 1) \geq 0 \quad ; \quad t \in \mathbb{R}_+$$

and

$$\psi_\alpha''(t) = \alpha\,(t^{\alpha-2} + t^{-\alpha-1}) \geq 0 \quad ; \quad t \in \mathbb{R}_+ \, .$$

This concludes the proof.

Acknowledgement: The authors thank Robert Boudreau for the computer production of Figure 1.

References

[1] Atkinson, C. and Mitchell, A. F. S., "Rao's distance measure", Sankhya (1980), in press.

[2] Good, I. J., "Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables", Ann. Math. Stat., Vol. 34, pp. 911-934, 1963.

[3] Havrda, M. E., and Charvát, F., "Quantification method of classification processes: Concept of structural α-entropy", Kybernetica, Vol. 3, pp. 30-35, 1967.

[4] Jeffreys, H., "An invariant form for the prior probability in estimation problems", Proc. Roy. Soc. London, Ser. A., Vol. 186, pp. 453-461, 1946.

[5] Kullback, S. and Leibler, R. A., "On information and sufficiency", Ann. Math. Statist., Vol. 22, pp. 79-86, 1951.

[6] Lewontin, R. C., "The apportionment of human diversity", Evolutionary Biology, Vol. 6, pp. 381-398, 1972.

[7] Rao, C. R., "Information and accuracy attainable in the estimation of statistical parameters", Bull. Calcutta Math. Soc., Vol. 37, pp. 81-91, 1945.

[8] Rao, C. R., "On the distance between two populations", Sankhya, Vol. 9, pp. 246-248, 1949.

[9] Rao, C. R., "Diversity and dissimilarity coefficients: a unified approach", University of Pittsburgh Tech. Rep. 80-10, 1980.

[10] Shannon, C. E, "A mathematical theory of communications", Bell System Tech. J., Vol. 27, pp. 379-423, 623-656, 1948.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER **AFOSR-TR- 80 - 1167** ✓	2. GOVT ACCESSION NO. AD-A093 123	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) On the Convexity of Some Divergence Measures Based on Entropy Functions		5. TYPE OF REPORT & PERIOD COVERED INTERIM
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) J. Burbea and C. Radhakrishna Rao		8. CONTRACT OR GRANT NUMBER(s) F49620-79-C-0161
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Pittsburgh Department of Mathematics & Statistics Pittsburgh, PA. 15260		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS AFOSR BOLLING AFB, WASHINGTON DC 20332		12. REPORT DATE September 1980
		13. NUMBER OF PAGES 26
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release: distribution unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Cross entropy, Divergence, Entropy, Jensen difference		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Three measures of divergence between vectors in a convex set of an n-dimensional real vector space have been defined in terms of certain types of entropy functions, and their convexity property studied. Among other results, a classification of the α-order entropies is obtained by the convexity of these measures. These results have applications to the measurement of diversity of a discrete probability distribution and divergence between two distributions.		

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

LME